

# METHOD AND SYSTEM FOR AUTOMATICALLY CREATING CROSSTALK-CORRECTED DATA OF A MICROARRAY

## TECHNICAL FIELD

5 This invention relates to methods and systems for creating crosstalk-corrected data of a microarray and, in particular, to methods and systems for automatically creating crosstalk-corrected data of a microarray utilizing calibration spots.

## BACKGROUND ART

10 Multifluorescence confocal imaging typically utilizes a multi-channel microarray scanner to obtain images of dye spots of a microarray. As illustrated in Figure 1, microarrays are created with fluorescently labeled DNA samples in a grid pattern consisting of rows 22 and columns 20 typically spread across a 1 by 3 inch glass microscope slide 24. Each spot 26 in the grid pattern 28 represents a separate DNA probe and constitutes a separate experiment. A plurality of such grid pattern  
15 comprises an array set 30. Reference or "target" DNA (or RNA) is spotted onto the glass slide 24 and chemically bonded to the surface. Fluorescently labeled "probe" DNA (or RNA) is introduced and allowed to hybridize with the target DNA. Excess probe DNA that does not bind is removed from the surface of the slide 24 in a subsequent washing process.

20 As illustrated in Figure 2, a confocal laser microarray scanner or microarray reader is commonly used to scan the microarray slide 24 to produce one image for each dye used by sequentially scanning the microarray with a laser of a proper wavelength for the particular dye. Each dye has a know excitation spectra as illustrated in Figure 3 and a known emission spectra as illustrated in Figure 4.  
25 The scanner includes a beam splitter 32 which reflects a laser beam 34 towards an objective lens 36 which, in turn, focuses the beam at the surface of slide 24 to cause fluorescent spherical emission. A portion of the emission travels back through the lens 36 and the beam splitter 32. After traveling through the beam splitter 32, the

fluorescence beam is reflected by a mirror 38, travels through an emission filter 40, a focusing detector lens 42 and a central pinhole 44. After traveling through the central pinhole 44, the fluorescence beam is detected by a detector, all in a conventional fashion.

5           The intent of a microarray experiment is to determine the concentrations of each DNA sample at each of the spot locations on the microarray. Further data analysis of the brightness values are typically done to produce a ratio of one dye's brightness to any or all of the other dyes on the microarray. An application of the microarray experiment is in gene expression experiments. Higher  
10   brightness values are a function of higher concentrations of DNA. With a microarray, a researcher can determine the amount a gene is expressed under different environmental conditions.

15           To be accurate, the reader must be able to quantitate the brightness of each microarray spot for each labeled DNA sample used in the experiment. To do this the reader must filter the emissions from any and all other fluorescent samples. The concentration of the DNA is a function of the brightness of the emission when excited by a laser of the proper wavelength. It becomes difficult to differentiate between the emissions of different dyes when the emission spectra of a dye overlaps with another. Furthermore, the brightness produced from the  
20   emission of one dye could be contaminated by emissions from another dye. This contamination of the brightness values is commonly known as crosstalk.

25           Microarray readers have been designed to simultaneously scan more than two dyes using lasers with the proper wavelength. In this type of experiment, multiple samples of DNA are hybridized onto the microarray, each with a different fluorescent label. Crosstalk contamination is equally likely as in the two dye experiments and can even be more troublesome when dyes with close emission spectra are placed on the same microarray.

U.S. Patent Nos. 5,804,386 and 5,814,454 disclose sets of labeled energy transfer fluorescent primers and their use in multi-component analysis.

U.S. Patent No. 5,821,993 discloses a method and system for automatically calibrating a color camera in a machine vision system.

The paper by Schena, M., et al., (1995) "Quantitative Monitoring of Gene Expression Patterns With a Complementary DNA Microarray", Science 270; 5 467-469 is also related to the present invention.

### DISCLOSURE OF INVENTION

669720 0054550  
10 An object of the present invention is to provide a method and system for creating crosstalk-corrected data of a microarray wherein a sequence of algebraic operations are used to obtain correction factors which, in turn, are used to correct for crosstalk between two or more dyes in a multi-channel imager such as a microarray scanner.

Another object of the present invention is to provide a method and system for creating crosstalk-corrected data of a microarray by utilizing calibration spots on a microarray sample substrate.

15 In carrying out the above objects and other objects of the present invention, a method is provided for automatically creating crosstalk-corrected data of a microarray. The method includes providing a microarray substrate having calibration dye spots. Each of the calibration dye spots comprises a single pure dye. The method also includes, for each of the calibration dye spots, generating a dye  
20 image containing at least one of the calibration dye spots for each of a plurality of output channels and also, for each of the calibration dye spots, measuring an output of each of the output channels to obtain output measurements. The method further includes computing a set of correction factors from the output measurements and applying the set of correction factors to data obtained from microarray images  
25 containing spots having dyes with excitation or emission spectra to obtain crosstalk-corrected data.

Preferably, the step of generating includes the step of imaging the calibration dye spots to produce a dye image for each calibration dye spot.

Preferably, the substrate is a glass slide.

Also, preferably, each of the channels is optimized for a different dye  
5 and the step of generating is performed by an imager such as a microarray scanner or a camera.

Preferably, each of the dyes is a fluorescent dye.

Preferably, the step of computing includes the step of computing  
10 crosstalk ratios based on spot brightness values for each of the calibration dye spots on each of the output channels.

Preferably, the number of calibration dye spots is more than or equal to the number of dyes.

The calibration dye spots may be hybridized target DNA and  
fluorescently labeled probe DNA.

15 Still further in carrying out the above objects and other objects of the present invention, a system is provided for carrying out the above method steps.

In the method and system of the present invention, crosstalk  
correction requires the availability and use of calibration spots on the microarray.  
As illustrated in Figure 5, these calibration spots should be composed of the highest  
20 concentration of each single probe or dye that could be obtained by the microarray process being utilized. By measuring the crosstalk between the calibration spots, one can obtain all of the information that is needed to correct for crosstalk in all spots of the microarray without explicit knowledge of the dyes' excitation or emission characteristics.

In the case of 'n' samples on the microarray experiment with each DNA sample labeled (*i.e.*, typically 1000-5000 spots but only 2-4 dyes), the number of crosstalk calibration spots is typically greater than or equal to the number of dyes used. More calibration spots can be used to better tolerate experimental abnormalities. In the case of additional calibration spots, all the spots of an identical dye can be averaged together. The dyes used to create the calibration spots should also be the same as were used to label the DNA samples as illustrated in Figure 6.

The above objects and other objects, features and advantages of the present invention are readily apparent from the following detailed description of the best mode for carrying out the invention when taken in connection with the accompanying drawings.

### BRIEF DESCRIPTION OF DRAWINGS

FIGURE 1 is a top plan schematic view illustrating a spot, an array and an array set on a glass slide;

FIGURE 2 is a schematic view of a confocal laser reader used to generate digital images;

FIGURE 3 illustrates graphs of sample excitation spectra;

FIGURE 4 illustrates graphs of sample emission spectra;

FIGURE 5 is a schematic view of calibration spots with two dyes;

FIGURE 6 is a schematic view of calibration spots with 'n' dyes;

FIGURE 7 is a schematic diagram illustrating a preferred hardware configuration on which the computational portion of the method of the present invention can be implemented; and

FIGURE 8 is a schematic view of a system in which the present invention can be utilized.

### BEST MODE FOR CARRYING OUT THE INVENTION

Referring now to the drawing figures, there is illustrated in Figure 7 a workstation on which the computational portion of the method and system of the present invention can be implemented. However, other configurations are possible. The hardware illustrated in Figure 7 includes a monitor 10 such as a single SVGA display, a keyboard 12, a pointing device such as a mouse 14, a magnetic storage device 16, and a chassis 18 including a CPU and random access memory. The monitor 10 may be a touch screen monitor used in addition to standard keyboard/mouse interaction. In a preferred embodiment, the chassis 18 is a Pentium-based IBM compatible PC or other PC having at least 32 megabytes of RAM and at least 12 megabytes of hard disk space. The workstation typically includes a Windows NT, graphical user interface as well as an Ethernet 10 Base-T high speed Lan network interface.

One or more images are obtained by a user from the microarray reader or scanner of Figures 2 and 8. The scanner is controlled by a scanner control computer 50 which, in turn, is also networked to a quantitation computer 52.

#### Calibration in the two channel microarray experiment

Assume that the user has provided two microarray spots for calibration as illustrated in Figure 5. Further assume two-color, two-channel scanning, with the microarray reader's channels balanced on these two calibration spots. The two dyes are called "Dye A" and "Dye B," and the instrument channels are called "Channel 1" and "Channel 2." Channel 1 is optimized for Dye A, and Channel 2 is optimized for Dye B. These calibration spots should contain "pure" dye, or more precisely, the maximum labeled-DNA concentration associated with 100% gene expression.

These calibration spots are referred to as Cal Spot A and Cal Spot B. Before scanning them, the channels of the reader are balanced to produce roughly equivalent brightness values on a spot other than a Cal Spot. Crosstalk is a relatively small (2-5%) signal in the opposite channel. The two dots are scanned,  
 5 both dots on both channels, and the scan data analyzed to produce spot brightness values. The four resulting data values are named as follows:

CalBrightA1 = spot brightness value of Cal Spot A scanned on Channel 1  
 CalBrightA2 = spot brightness value of Cal Spot A scanned on Channel 2  
 CalBrightB1 = spot brightness value of Cal Spot B scanned on Channel 1  
 10 CalBrightB2 = spot brightness value of Cal Spot B scanned on Channel 2

Crosstalk ratios are defined as follows:

Crosstalk A =  $\text{CalBrightA2} / \text{CalBrightA1}$

Crosstalk B =  $\text{CalBrightB1} / \text{CalBright B2}$

These two measured Crosstalk values (which are each a fraction less  
 15 than 1) are stored for use in correcting values on all of the other dots on the array.

### **Correction in the two channel microarray experiment**

The other dots in the array have random combinations of Dye A and Dye B in unknown ratios. Each dot is scanned on both Channel 1 and Channel 2, and those two raw brightness values are corrected for crosstalk. The first-order  
 20 method for doing that is as follows.

Define more terms: "Brightness" is the measured intensity value for a spot from a particular instrument channel. "Signal" (1 and 2) is the portion of brightness (presumably the large majority) which is from the target dye (e.g., not crosstalk). "Signal" (1 and 2) is the answer that is sought.

25                    Unknowns:    Signal 1 =  $S_1$

$$\text{Signal 2} = S_2$$

Knowns: What is measured on each spot

$$\text{Brightness 1} = B_1$$

$$\text{Brightness 2} = B_2$$

5 From the two channel calibration section

$$\text{Crosstalk}_{12} = \alpha_{12}$$

$$\text{Crosstalk}_{21} = \alpha_{21}$$

Signal n for each spot on the array can then be determined by the following equations:

$$10 \quad B_1 = S_1 + S_2 \alpha_{12}$$

$$B_2 = S_2 + S_1 \alpha_{21}$$

or, solving for Signal:

$$S_1 = (B_1 - (\alpha_{12} \times B_2)) / (1 - (\alpha_{12} \times \alpha_{21}))$$

$$S_2 = (B_2 - (\alpha_{21} \times B_1)) / (1 - (\alpha_{12} \times \alpha_{21}))$$

## 15 Calibration and correction in the 'n' channel microarray experiment

Scanners with 3, 4, or more channels are perhaps even more likely to suffer from crosstalk than 2-channel instruments. Correction for this is accomplished using the same calibration spot technique, and the measurement of the crosstalk contribution of all of the combinations of excitation wavelengths and dyes.

20 To generalize some definitions of terms:

$\alpha_{xy}$  = measured and calculated crosstalk ratio of Dye Y into the Dye X channel

$S_x$  = Signal from Dye X (which one is seeking)

25  $B_x$  = Measured brightness of an arbitrary spot on the Dye X channel



Then, for the 3-channel case, the equations are as follows:

$$B_1 = S_1 + S_2\alpha_{12} + S_3\alpha_{13}$$

$$B_2 = S_1\alpha_{21} + S_2 + S_3\alpha_{23}$$

$$B_3 = S_1\alpha_{31} + S_2\alpha_{32} + S_3$$

5 which, in matrix form looks like:

$$[B_x] = [S_x][A] \text{ where } A = \begin{bmatrix} 1 & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & 1 & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & 1 \end{bmatrix}$$

$$\text{DET } A = 1 - \alpha_{12}\alpha_{21} - \alpha_{13}\alpha_{31} - \alpha_{23}\alpha_{32} + \alpha_{13}\alpha_{21}\alpha_{32} + \alpha_{12}\alpha_{23}\alpha_{31}$$

$$S_1 = \frac{B_1(1 - \alpha_{23}\alpha_{32}) - B_2(\alpha_{12} - \alpha_{13}\alpha_{32}) + B_3(\alpha_{12}\alpha_{23} - \alpha_{13})}{\text{DET } A}$$

$$S_2 = \frac{-B_1(\alpha_{21} - \alpha_{31}\alpha_{23}) + B_2(1 - \alpha_{31}\alpha_{13}) - B_3(\alpha_{23} - \alpha_{21}\alpha_{13})}{\text{DET } A}$$

$$10 \quad S_3 = \frac{B_1(\alpha_{21}\alpha_{32} - \alpha_{31}) - B_2(\alpha_{32} - \alpha_{12}\alpha_{31}) + B_3(1 - \alpha_{12}\alpha_{21})}{\text{DET } A}$$

The expansion of this matrix from 3 x 3 to 4 x 4 (or n x n) is straightforward.

15 While embodiments of the invention have been illustrated and described, it is not intended that these embodiments illustrate and describe all possible forms of the invention. Rather, the words used in the specification are words of description rather than limitation, and it is understood that various changes may be made without departing from the spirit and scope of the invention.